



Navigating the Sea of Options

File formats and Naming for Long Term Maintenance and Access

Kelly Eubank

North Carolina State Archives

November 3, 2011

Public Records Law

General Statute § 132-1 Public Records Act

a) "Public record" or "public records" shall mean all documents, papers, letters, maps, books, photographs, films, sound recordings, magnetic or other tapes, electronic data-processing records, artifacts, or other documentary material, regardless of physical form or characteristics, made or received pursuant to law or ordinance in connection with the transaction of public business by any agency of North Carolina government or its subdivisions.

G.S. 121

Archives and History Act

- 121-4(2)
 - Assigns records management responsibility
- 121-5(b)
 - Regulates the destruction of public records
- 121-5(c)
 - Directs DCR to assist governments and agencies
 - State Archival Agency Designated. – The Department of Cultural Resources shall be the official archival agency of the State of North Carolina

How is digital different?

- Cannot view without manipulation,
- Format dependency-hardware & software,
- Logical file names,
- have to be managed over time to ensure access.

What can you do?

- **Format awareness**

File formats—Is the data format open or proprietary?

Open standards and APIs improve interoperability



- **Vendor support lifecycle and documentation**

What formats will vendor use? Can you get your stuff back and be able to use it?

Beware of both trendy and aging formats

Is the data stored in a single file/container or requires multiple files?

What to look for

- Openness/Transparency—is format documented
- Adopted as a standard (a must for anything longer than 10 years)
- Stability
- Dependencies/Interoperability—hardware, software



CHOOSING GOOD FILE FORMATS

Importance of File Formats and Management



Name ▲	Size	Type
 1992_NC_Congressional_Districts.dbf	3 KB	DBF File
 1992_NC_Congressional_Districts.prj	1 KB	PRJ File
 1992_NC_Congressional_Districts.sbn	1 KB	SBN File
 1992_NC_Congressional_Districts.sbx	1 KB	Adobe Illustrator Ts...
 1992_NC_Congressional_Districts.shp	2,020 KB	SHP File
 1992_NC_Congressional_Districts.shp.xml	31 KB	XML Document
 1992_NC_Congressional_Districts.shx	1 KB	SHX File

Recommended preservation formats?



**Word Processing Files: PDF (.pdf)
TEXT (.txt) *, or RTF (.rtf)**



**Images: TIFF (.tif) or JPEG2000
(.jp2)***



**Audio: MPEG-2 (.mpg, .mpeg,
.mp2) or MPEG-4**



Video: WAV (.wav) w/BWF header

How do you know what it is without having to open it?

FILE NAMING

TF_Request_Form_DCR_v1_20090610.xls
Travel Request Form DCR v1 200906...
Section; Division: Government Records Branch: Archiv
Destination: Raleigh, NC to Albany, NY--Best Practices
Travel Dates from 09/01/2009 to G
(include times)
Travel Request Form Completed by: Kelly Eubank

Purpose of Travel:

Kelly Eubank will be attending the Best Practices Exchange 2009 Ar for the GeoMAPP project regarding preservation and providing acce: acquiring, preserving, and accessing complicated geospatial dataset State Access and Preservation Partnership to which Mary's time is d Payment is requested for a fee to change airfare to/from Albany, NY through the special revenue account 4602 2165. This request is an airport.

From: [REDACTED]
To: Kelly Eubank
Cc:
Subject: Travel Request

File sent via e-mail

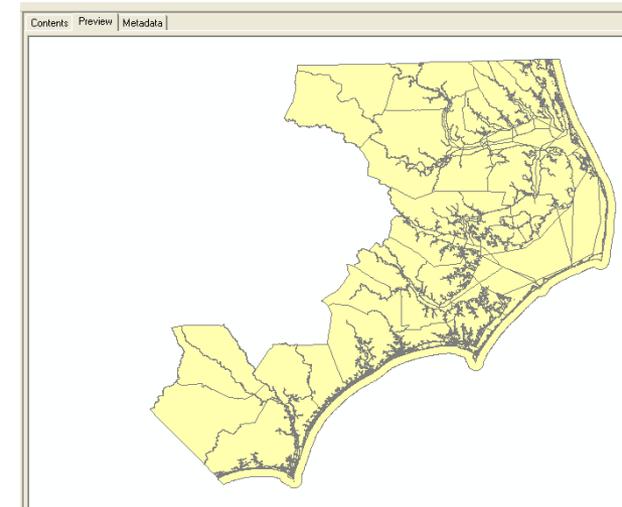
Message | 2009_08_31_15_20_30.pdf (253 KB)

This attachment is for your records.

Best Practices

- File naming
 - Descriptive title
 - Wake_Parcels_2006
 - Shellfish_Growing_Areas_2009
- Attributes
 - Logical Name
 - Metadata description

Name	Size	Type
sga.dbf	2,502 KB	DBF File
sga.prj	1 KB	PRJ File
sga.sbn	86 KB	SBN File
sga.sbx	4 KB	SBX File
sga.shp	49,668 KB	SHP File
sga.shx	68 KB	SHX File
sga.txt	23 KB	Text Document



Name	Size	Type
Hurricane_Storm_Surge_Slow_1999_09.dbf	632 KB	DBF File
Hurricane_Storm_Surge_Slow_1999_09.prj	1 KB	PRJ File
Hurricane_Storm_Surge_Slow_1999_09.sbn	62 KB	SBN File
Hurricane_Storm_Surge_Slow_1999_09.sbx	3 KB	SBX File
Hurricane_Storm_Surge_Slow_1999_09.shp	67,289 KB	SHP File
Hurricane_Storm_Surge_Slow_1999_09.shp.xml	32 KB	XML Document
Hurricane_Storm_Surge_Slow_1999_09.shx	51 KB	SHX File

Example of Nirvana

Name ▲	Size	Type	Date Modified
 71_Emergency_Ice&Snow_05...	96 KB	Adobe Acrobat Doc...	6/17/2005 2:03 PM
 72_Terminate EO45&71.pdf	83 KB	Adobe Acrobat Doc...	6/17/2005 2:05 PM
 73_NC Motorsports_Adv_Co...	139 KB	Adobe Acrobat Doc...	6/17/2005 2:06 PM
 74_Replace EO75&118.pdf	389 KB	Adobe Acrobat Doc...	6/17/2005 2:09 PM
 75_2005_CHAF_Rules.pdf	151 KB	Adobe Acrobat Doc...	6/13/2005 4:24 PM
 76_Extend_EO1_BdofEthics.pdf	42 KB	Adobe Acrobat Doc...	6/16/2005 10:32 AM
 77_TeacherAdvisoryCommitte...	90 KB	Adobe Acrobat Doc...	6/30/2005 5:28 PM
 78_DisasterTowns of Carolina...	145 KB	Adobe Acrobat Doc...	7/11/2005 4:51 PM
 79_NC Film Council.pdf	107 KB	Adobe Acrobat Doc...	7/18/2005 11:02 AM
 80_Leandro_REV.pdf	167 KB	Adobe Acrobat Doc...	7/21/2005 4:15 PM
 81_Katrina_Relief.pdf	148 KB	Adobe Acrobat Doc...	9/1/2005 2:01 PM
 82_Katrina_Proclamation.pdf	92 KB	Adobe Acrobat Doc...	9/7/2005 1:33 PM
 83_JuvenileJustice PlanCommi...	106 KB	Adobe Acrobat Doc...	9/9/2005 11:48 AM
 84_EmergencyResponseCom...	100 KB	Adobe Acrobat Doc...	9/9/2005 11:49 AM
 85_HispanicLatinoAdvCouncil...	95 KB	Adobe Acrobat Doc...	9/9/2005 11:50 AM
 86_FlexibleBenefitsProg.pdf	119 KB	Adobe Acrobat Doc...	9/9/2005 11:51 AM
 87_Amend81_KatrinaRelief.pdf	163 KB	Adobe Acrobat Doc...	9/8/2005 3:22 PM
 88_Hurricane_Ophelia.pdf	86 KB	Adobe Acrobat Doc...	9/12/2005 4:19 PM

Metadata

Information that Matters

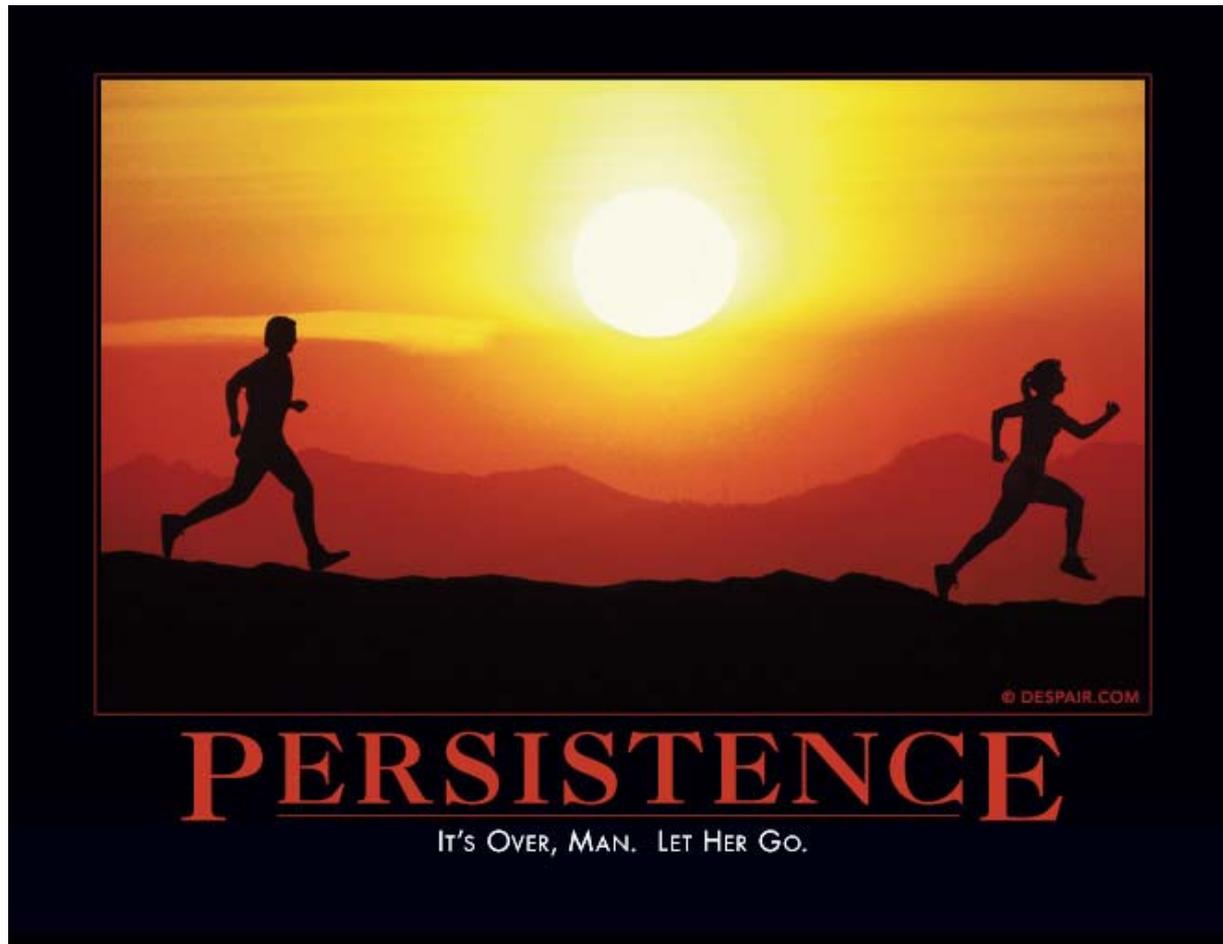
- into existing DCR responsibilities, services, existing services, have accomplished the following:
- Offered workshops and tutorials to
 - Tested and implemented tools for the electronic files
 - Developed guidelines to support digital
 - Worked collaboratively within DCR

Substantive metadata is information stored about the content of a file, such as the "track changes" feature in documents.

The image shows a Windows 'Properties' dialog box for a file named 'bluegill.jpg'. The 'General' tab is selected, showing file details such as 'Type of file: JPEG Image', 'Location: D:\ANIMAL', and 'Size: 6.19 MB (6)'. A red arrow points from the 'Summary' tab area to an Excel spreadsheet. The spreadsheet shows a formula bar with '=SUM(G12:G16)' and a grid of cells. A red arrow points from cell G12 (containing '\$0.00') to cell G16 (containing '\$0.00').

Embedded metadata is information that is manually created by the user

You CAN do it!



Best Practices Recap

- File naming
 - Descriptive title
 - Wake_Parcel_2006
- Formats
 - Use common standard formats
- Metadata– Do IT!
 - Comply with existing Standards
- You are not alone, we are here to help!
 - www.records.ncdcr.gov



Resources

[Library and Archives of Canada Local Digital Format Registry](http://www.collectionscanada.gc.ca/digital-initiatives/012018-2210-e.html#c)

<http://www.collectionscanada.gc.ca/digital-initiatives/012018-2210-e.html#c>

Recommended Digital File Format—North Carolina State Archives

http://www.records.ncdcr.gov/erecords/FileFormatTable-Access-March_2007.pdf

PADI file formats bibliography

<http://www.nla.gov.au/padi/topics/612.html>

North Carolina Digital Preservation Education Website

www.digitalpreservation.ncdcr.gov

PADI file formats bibliography

<http://www.nla.gov.au/padi/topics/612.html>

- Contact Information
- Kelly Eubank
- 919-807-7355
- Kelly.eubank@ncdcr.gov
- www.records.ncdcr.gov
- www.digitalpreservation.ncdcr.gov

